Docket No. RSW920000182US1

ABSTRACT OF THE DISCLOSURE

METHOD AND SYSTEM FOR SAMPLE DATA SELECTION TO TEST AND TRAIN PREDICTIVE ALGORITHMS OF CUSTOMER BEHAVIOR

5

A method and system for sample data selection to test and train predictive algorithm of customer behavior are provided. The method and system generate frequency distributions of a customer database data set, training
10   data set and testing data set and compare the frequency distributions of geographical characteristics to determine if there are discrepancies. If the discrepancies are above a predetermined tolerance, one or more of the data sets may not be representative of the
15   customer database taking into account geographical influences on customer behavior. Thus, recommendations for improving the training data set and/or testing data set are then provided such that the data set is more representative of the customer database. In this way,
20   "nuggeting" of customers is accounted for in the training and/or testing data sets.